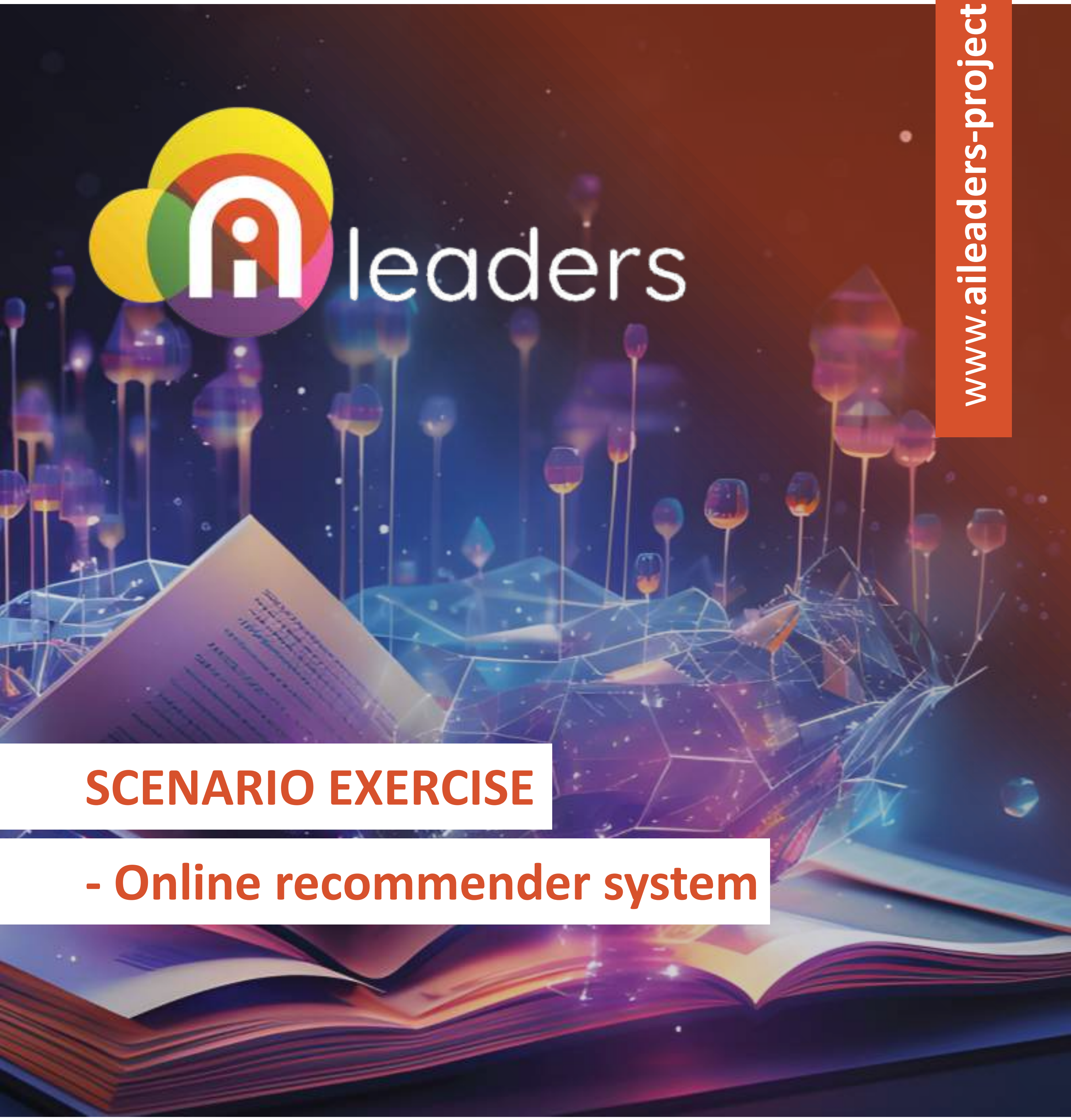
www.aileaders-project.eu



AI leaders

SCENARIO EXERCISE

- Online recommender system

SCENARIO EXERCISE
- Online recommender system

01 Abstract

Type of OER

Scenario Exercise using Google Colab (riverML)

Goal/Purpose

This scenario-based exercise puts students in the role of a machine learning (ML) practitioner tasked with analysing the fairness of online recommender system. It explores how continuous adaptation to user behaviour can amplify existing biases over time. Students will be challenged to identify bias and evaluate mitigation strategies.

Expected Learning Outcomes:

The student will be able to:

- Identify potential sources of bias in ML-generated recommendations;
- Compare fairness metrics such as demographic parity and exposure equity;
- Understand how feedback loops affect fairness over time in online recommender systems.

Keywords:

- Machine Learning
- Recommendation
- Biases
- Fairness

Suggested Methodological Approach:

Problem-Based Learning.

02 Introduction

In the context of increasingly personalised digital experiences, AI-driven recommender systems play a pivotal role in shaping consumer behaviour, influencing purchasing decisions, and determining the visibility of products and content across user segments.

While these systems offer considerable advantages for marketing and business optimisation, they also raise pressing ethical and regulatory concerns—particularly regarding fairness, transparency, and potential discrimination (Akter et al., 2022; Bozdag, 2013).

Recommender systems are not merely algorithmic artefacts but complex sociotechnical constructs, influenced by human decisions at every stage—from data collection and model design to output interpretation and deployment (Bozdag, 2013). Biases can emerge from multiple sources: human prejudices, technical limitations, or contextual misalignments. In marketing environments especially, these biases can interact and compound, shaping not only system performance but also user experiences and societal outcomes (Akter et al., 2022).

One significant concern lies in the iterative nature of recommender systems, which depend on user feedback to retrain and refine predictive models. This feedback loop may reinforce existing input-output imbalances, leading to persistent disparities over time (Sun et al., 2020). As models continuously incorporate new user data and preferences, their fairness and accuracy are subject to dynamic change. If left unchecked, this can result in overfitting to dominant user behaviours, exacerbating the under-representation of minority groups and restricting content diversity—a phenomenon associated with algorithmic gatekeeping and the rise of filter bubbles (Harambam et al., 2018).

Motivated by these challenges, this scenario-based exercise invites students to explore the ethical dimensions of recommender systems using the **MovieLens 100K dataset** as a testbed.

Through hands-on experimentation, students examine how algorithmic feedback loops may unintentionally amplify inequalities, particularly by skewing content visibility and favouring majority preferences. They are encouraged to investigate both technical and ethical aspects of these systems, bridging theoretical insight with practical analysis. This pedagogical approach fosters critical awareness of the risks associated with unethical personalisation—including loss of consumer trust, reputational harm, and diminished access to opportunities for marginalised users. The exercise highlights that fairness in AI is not a static evaluation, but a dynamic, ongoing responsibility. As user preferences evolve and content shifts, so too must models adapt to maintain balanced performance across all user groups. Ultimately, this learning experience equips students with the tools to design AI systems that are not only effective and efficient, but also equitable, transparent, and socially responsible.

Fairness Metrics

01 **Absolute estimation error** - measures the absolute error between the predicted rating and the true rating given by user u to item I, defined by:

$$\varepsilon_a(u,i) = |\,\hat{r}_{u,i} - r_{u,i}\,|$$

02 **Overestimation error** - measures how much the prediction overestimates the true rating, defined by:

$$\varepsilon_o(u,i) = \max(\hat{r}_{u,i} - r_{u,i}, 0)$$

03 **Underestimation error** - measures how much the prediction underestimates the true rating, defined by:

$$\varepsilon_u(u,i) = \max(r_{u,i} - \hat{r}_{u,i}, 0)$$

04 **Value estimation error** - measures the signed error between the predicted rating and the true rating, defined by:

$$\varepsilon_v(u,i) = \hat{r}_{u,i} - r_{u,i}$$

03 Tools Presentation

This exercise leverages the **MovieLens 100K dataset**, a widely-used benchmark in recommender system research. The dataset includes 100,000 ratings, 943 users, 1,682 movies, along with user demographic data (e.g., age, gender, occupation).

While useful for algorithm testing, the dataset contains significant fairness challenges:

01 **Imbalanced representation across demographic groups;**

02 **Skewed rating distributions;**

03 **Over-representation of popular items.**

These characteristics make MovieLens 100K ideal for exploring bias and fairness in recommendation algorithms.

RiverML[1] is an open-source Python library specifically designed for machine learning on data streams. It provides tools for building, training, and evaluating models that learn incrementally from data arriving in real time, rather than in batch. RiverML supports both supervised and unsupervised learning algorithms, offering features like adaptive learning, concept drift detection, and stream preprocessing. Its modular design makes it easy to combine models, preprocessing steps, and evaluation strategies, making it a powerful choice for online learning applications such as fraud detection, recommendation systems, and sensor data analysis. The fairness metrics available in the library were implemented by the FEP team, contributing to the ethical evaluation of streaming models.

This tool enables the execution of fairness evaluation pipelines and visual comparison of recommendation performance across groups and time

[1] Riverml Tutorial - https://riverml.xyz/latest/introduction/installation/

04 Hands-on Activities

Activity:

01 Access the provided Jupyter Notebook at https://tinyurl.com/5n8vvpx9.

02 Execute all cells in the notebook (Click on Runtime Tab followed by Run All)

03 Analyse and compare fairness metrics across different user groups (e.g., by gender, age).

04 Observe how fairness indicators evolve as more user interactions are collected

Discussion Prompts:

01 Which user groups are systematically favoured or disadvantaged?

02 How does recommendation quality change over time?

03 What trade-offs emerge when trying to balance engagement with fairness?

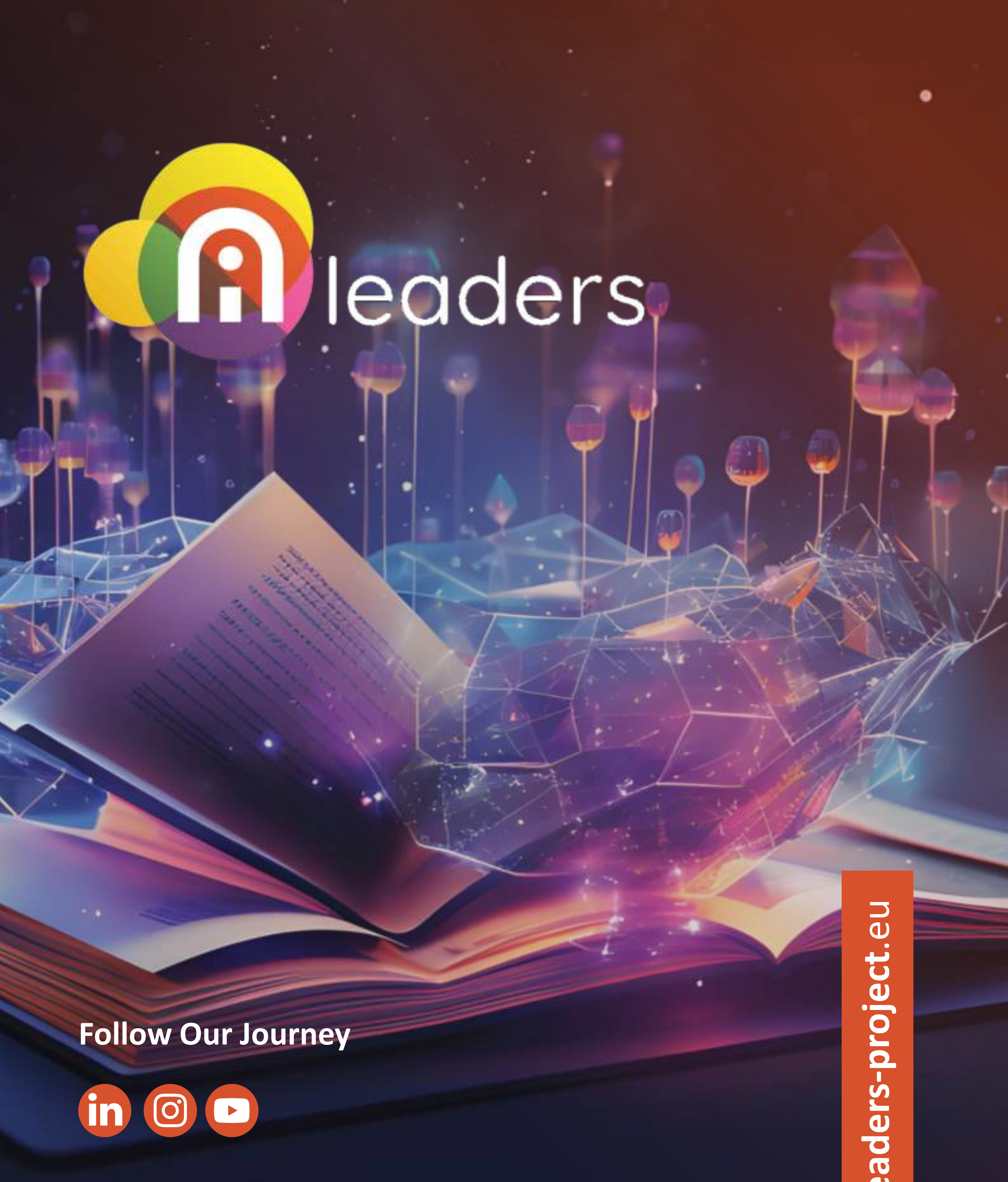
05 Conclusion

This exercise illustrates the dynamic nature of fairness in online recommendation systems.

While initial system performance may appear unbiased, long-term operation can lead to significant disparities due to feedback loops and popularity bias. These imbalances affect user experience and trust, particularly for underrepresented or minority user groups. By applying fairness metrics and analysing the evolution of bias in real time, students gain practical insight into ethical ML deployment. They also explore the limitations of static evaluation and the importance of continuous fairness monitoring in adaptive systems. The exercise underlines the tension between optimising for engagement and ensuring equitable treatment - a key challenge for ethical AI design.

06 References

- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. Ethics and Information Technology, 15(3), 209–227.
- Akter, S., Michael, K., Bandara, R., Wamba, S.F., Foropon, C., & Papadopoulos, T. (2022). Algorithmic bias in machine learning–based marketing models: A dynamic capability view. Journal of Business Research.
- Harambam, J., Helberger, N., & van Hoboken, J. (2018). Democratizing algorithmic news recommenders: How to materialize voice in a technologically saturated media ecosystem. Big Data & Society.
- Sun, T., Gautam, K., & Joachims, T. (2020). Debiasing Learning from User Interaction. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.



Follow Our Journey

www.aileaders-project.eu

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.